

ACE Scoring for 2005

- **Scorer** overview
- Pilot annotation scores

The Scoring Method

- The **scorer** scores the performance of a system by computing the “value” of the system’s output using a three-step process:
 1. The value of each system output element is computed for all possible sys-ref mappings, including misses (sys doesn’t match any ref) and false alarms (ref doesn’t match any sys).
 2. An optimum association (one-to-one mapping) of sys elements to ref elements is found so that the resulting bottom-line score is maximized.
 3. The bottom-line score is computed, along with a myriad of diagnostic information, including various attribute-conditioned performance statistics and various attribute confusion matrices.

“Cross-document” = “Real World”

- The **scorer** scores the value of inferences made about things in the “real world”.
 - To be valuable these inferences must represent real-world elements that exist separate and apart from the document(s) that mention them.
 - Real world identity is represented in the **apf** file format by means of a globally unique ID that is assigned to each (unique) real-world element.
 - The **scorer** always performs real-world (i.e., “cross-doc”) scoring.
 - If only document-level scoring is desired, then the coreference links must be severed in the reference data (by creating unique element ID’s in each document).

The EDR Scoring Formula

$$EDR_Value_{sys} = \sum_i value_of_sys_entity_i$$

$$value_of_sys_entity = Entity_Value(sys_entity) \cdot \sum_m Mention_Value(sys_mention_m)$$

$$Entity_Value = \left\{ \begin{array}{l} \min \left(\frac{ETypeValue(sys) \cdot EClassValue(sys)}{ETypeValue(ref_{sys}) \cdot EClassValue(ref_{sys})} \right) \cdot (W_{Eerr-type} \cdot W_{Eerr-subtype} \cdot W_{Eerr-class}) \text{ when mapped} \\ ETypeValue(sys) \cdot EClassValue(sys) \cdot (W_{E-FA}) \text{ when entity not mapped} \end{array} \right\}$$

$$Mention_Value = \left\{ \begin{array}{l} \min \left(\frac{MTypeValue(sys)}{MTypeValue(ref_{sys})} \right) \cdot (W_{Merr-type} \cdot W_{Merr-role} \cdot W_{Merr-style}) \text{ when mapped} \\ - MTypeValue(sys) \cdot (W_{M-FA} \cdot W_{M-CR}) \text{ when mention not mapped} \end{array} \right\}$$

The QDR Scoring Formula

$$QDR_Value_{sys} = \sum_i value_of_sys_quantity_i$$

$$value_of_sys_quantity = Quantity_Value(sys_quantity) \cdot \sum_m Mention_Value(sys_mention_m)$$

$$Quantity_Value = \begin{cases} \min\left(\frac{QTypeValue(sys)}{QTypeValue(ref_{sys})}\right) \cdot (W_{Qerr-type} \cdot W_{Qerr-subtype}) & \text{when mapped} \\ WTypeValue(sys) \cdot (W_{Q-FA}) & \text{when quantity not mapped} \end{cases}$$

$$Mention_Value = \begin{cases} 1 & \text{when mapped} \\ -W_{Q-FA} & \text{when mention not mapped} \end{cases}$$

The RDR Scoring Formula

$$RDR_Value_{sys} = \sum_i value_of_sys_relation_i$$

$$value_of_sys_relation = Relation_Value(sys_relation) \cdot \sum_a Argument_Value(sys_argument_a)$$

$$Relation_Value = \begin{cases} \min\left(\frac{RTypeValue(sys)}{RTypeValue(ref_{sys})}\right) \cdot (W_{Rerr-type} \cdot W_{Rerr-subtype}) & \text{when mapped} \\ RTypeValue(sys) \cdot (W_{R-FA}) & \text{when relation not mapped} \end{cases}$$

$$Argument_Value = Element_Value(sys)$$

The VDR Scoring Formula

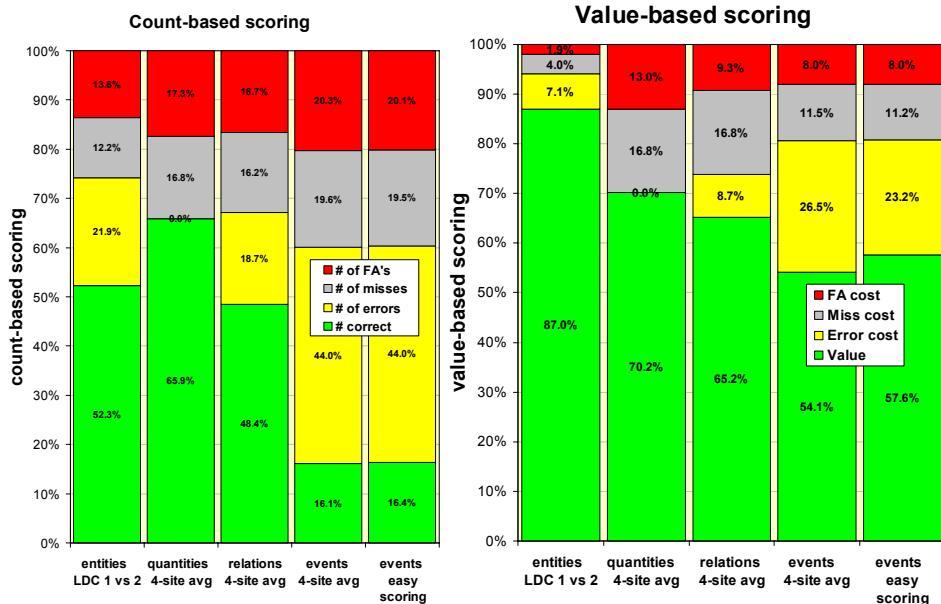
$$VDR_Value_{sys} = \sum_i value_of_sys_event_i$$

$$value_of_sys_event = Event_Value(sys_event) \cdot \sum_a Argument_Value(sys_argument_a)$$

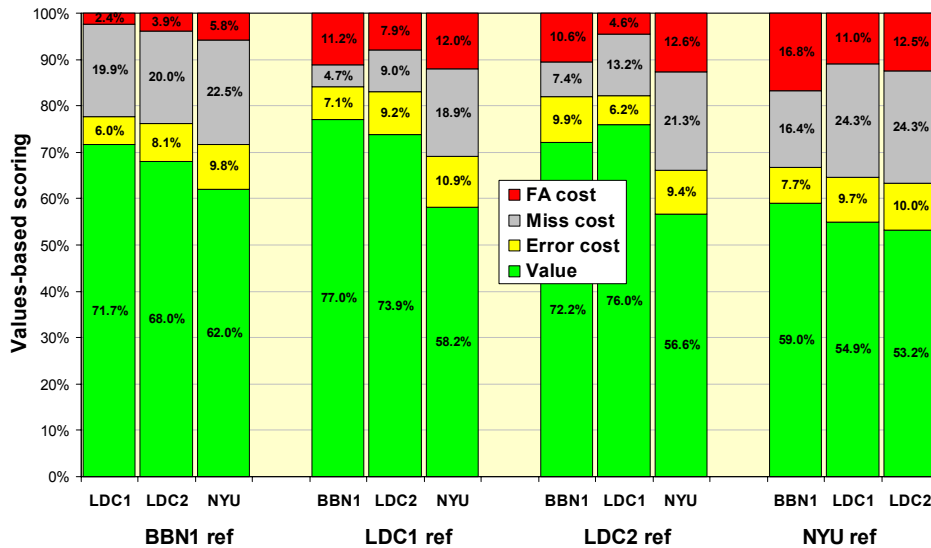
$$Event_Value = \begin{cases} \min \left(VTypeValue(sys) \cdot VModeValue(sys), VTypeValue(ref_{sys}) \cdot VModeValue(ref_{sys}) \right) \cdot (W_{Verr-type} \cdot W_{Verr-mode}) & \text{when mapped} \\ VTypeValue(sys) \cdot (W_{V-FA}) & \text{when event not mapped} \end{cases}$$

$$Argument_Value = \begin{cases} Element_Value(sys) \cdot (W_{Aerr-role}) & \text{when mapped} \\ Element_Value(sys) \cdot (W_{A-FA}) & \text{when argument not mapped} \end{cases}$$

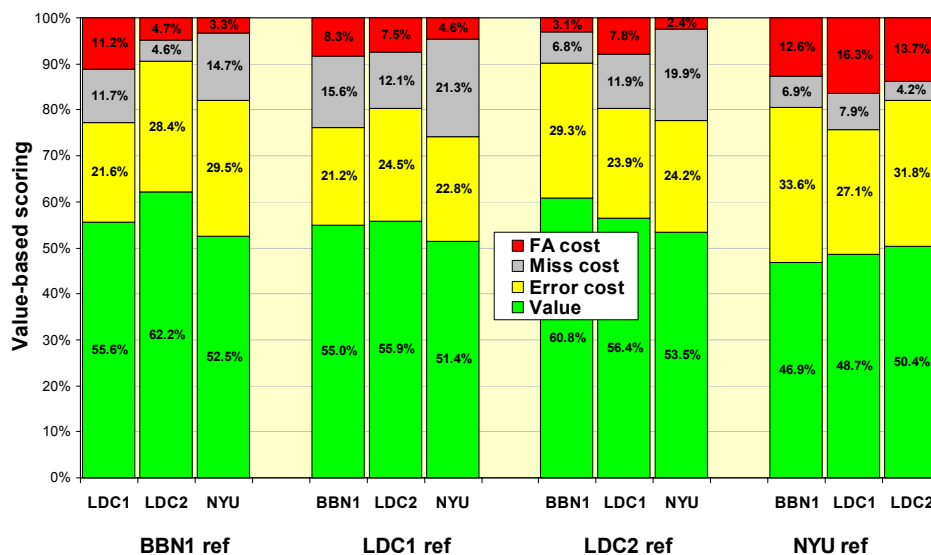
Pilot Corpus Annotation Scores



Interannotator Performance on Relation Annotation



Interannotator Performance on Event Annotation



Interannotator Performance on Quantity Annotation

